Intelligence 82 (2020) 101489

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Fluid reasoning is equivalent to relation processing

Jan Jastrzębski[a,*], Michał Ociepka[b], Adam Chuderski[a]

[a] Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Krakow, Poland
[b] Institute of Psychology, Jagiellonian University, Ingardena 6, 30-060 Krakow, Poland

ARTICLE INFO

Keywords:
Fluid reasoning
Relation processing
Working memory

ABSTRACT

Fluid reasoning (Gf)—the ability to reason abstractly—is typically measured using nonverbal inductive reasoning tests involving the discovery and application of complex rules. We tested whether Gf, as measured by such traditional assessments, can be equivalent to *relation processing* (a much simpler process of validating whether perceptually available stimuli satisfy the arguments of a single predefined relation—or not). Confirmatory factor analysis showed that the factor capturing variance shared by three relation processing tasks was statistically equivalent to the Gf factor loaded by three hallmark fluid reasoning tests. Moreover, the two factors shared most of their residual variance that could not be explained by working memory. The results imply that many complex operations typically associated with the Gf construct, such as rule discovery, rule integration, and drawing conclusions, may not be essential for Gf. Instead, fluid reasoning ability may be fully reflected in a much simpler ability to effectively validate single, predefined relations.

1. Introduction

The construct of fluid intelligence (Cattell, 1963), or fluid reasoning (Gf; Carroll, 1993), represents a broad cognitive ability related to the employment of such mental operations as “drawing inferences, concept formation, classification, generating and testing hypothesis, identifying relations, comprehending implications, problem solving, extrapolating, and transforming information” (McGrew, 2009, p. 5). The Gf factor is one of the most prominent psychological constructs, one that substantially predicts outcomes in multiple social domains—from academic and work success to health and longevity. As well, it strongly correlates with cognitive performance on various tasks, from mental speed and perception to long-term memory and verbal comprehension (Deary, 2012).

As “inductive and deductive reasoning are generally considered the hallmark indicators of Gf” (McGrew, 2009, p. 5), Gf is typically measured using nonverbal tests that require inductive reasoning and spatial visualization. Virtually all such tests involve the discovery of abstract rules and patterns that govern relatively complex figural stimuli and, further, the application of these rules/patterns to select the best matching response option. Probably the most influential Gf test is Raven's Advanced Progressive Matrices (RAPM; Raven, 1938; Raven, Raven, & Court, 1998). RAPM presents the variants of a 3 × 3 matrix of geometric patterns, of which a bottom right pattern is missing. The test requires inducing this pattern from the structure of the row- and column-wise changes (e.g., permutation, increase in number or value, application of logical operations such as AND, OR, and XOR, etc.) across the remaining eight patterns. Matrix problems are also included on another popular Gf test: Cattell's Culture Free Intelligence Test (CFT-3; Cattell, 1961). Its three other sections require pattern series completion, shape categorization, and understanding topological relations. In another kind of Gf test, figural analogies (Snow, Kyllonen, & Marshalek, 1984), the structure of geometric transformations must be identified between two patterns (A & B); then the same transformations must be applied to pattern C to infer the correct analog of B (A:B::C:? → D).

Charles Spearman, the founder of intelligence research, believed that RAPM and similar tests strongly involve two mental operations that are crucial to cognitive ability: so-called eduction of relations (abstracting a relation from the configuration of concrete elements) and eduction of correlates (predicting the configuration of the elements given the relation). In modern psychological terms, the operations stressed by Spearman can be described as identifying the valid relations when the values of its arguments are known (i.e., rule/relation discovery) and assigning the valid values to the arguments for the known relation (i.e., relation instantiation), respectively.

Recent research (Halford, Wilson, & Phillips, 1998; Oberauer, Süß, Wilhelm, & Wittmann, 2008) has supported the central role of relations in explaining individual differences in Gf. Moreover, multiple studies (see Gentner & Kurtz, 2005; Holyoak, 2012; Johnson-Laird, 2012) have

* Corresponding author.
E-mail addresses: jan.jastrzebski@uj.edu.pl (J. Jastrzębski), adam.chuderski@uj.edu.pl (A. Chuderski).

https://doi.org/10.1016/j.intell.2020.101489
Received 9 December 2019; Received in revised form 27 August 2020; Accepted 1 September 2020

suggested that traditionally distinguished kinds of reasoning, including deduction, induction, abduction, categorization, and analogies—all mentioned in the above definition of Gf—can be unified theoretically under the concept of *relational reasoning* (i.e., mental processing of relations in a way that preserves the valid assignment of values to the consecutive arguments of the relation, regardless of what exactly are the relation, values, and arguments). This property is called *structural consistency*; for instance, when *taller*[Tom, John] is transformed in the mind into *shorter*[X, Y], the value-argument assignments must remain (*shorter*[John, Tom]), and neither John's nor Tom's properties can change (see Doumas & Hummel, 2010). This may be a trivial task for a binary relation such as *taller* but becomes a highly demanding operation when the number of arguments increases (e.g., Birney & Bowman, 2009; Chuderski, 2014; Halford, Baker, McCredden, & Bain, 2005; Halford, Cowan, & Andrews, 2007).

However, typical fluid/relational reasoning tests require integrating multiple relations into multiple compound rules that must be inferred from relatively complex stimuli. They also involve elaborate management of subgoals (e.g., switching between inspecting the problem and analyzing the response options; Hayes, Petrov, & Sederberg, 2011; Jarosz & Wiley, 2012). All this entails a relatively long reasoning process that may be influenced by the use of various strategies (Jarosz, Raden, & Wiley, 2019) as well as by other factors. Thus, the exact role of processing relations for reasoning cannot readily be studied. To avoid such problems and more precisely examine processing relations, much simpler relational tasks have been developed, including the Latin Square Task (Birney & Bowman, 2009) and variants of the transitive reasoning task (Goodwin & Johnson-Laird, 2008).

Probably the most successful task in this vein is the relation monitoring task (Oberauer, Su, Wilhelm, & Wittman, 2003). In one variant of the measure, people observe a constantly changing matrix of symbols and decide whether all of the strings in one row, column, or diagonal line end with the same symbol. The task thus requires identifying relations among the symbols while imposing relatively low storage requirements (as all the information is available on-screen). Despite the task's simplicity, several of its variants have been shown to be surprisingly good predictors of Gf, above and beyond other tasks (Bateman, Thompson, & Birney, 2019; Chuderski, 2014; Oberauer et al., 2008).

The strong correlation of working memory (WM) capacity (WMC) with Gf (50–75% variance shared; see Kane, Hambrick, & Conway, 2005; Oberauer, Wilhelm, Schulze, & Süß, 2005) strongly implies that individual effectiveness of relation processing—a mental operation defined at a relatively abstract level—is determined by a limitation at the fine-grained level of WM. Although WM has received many definitions and models, generally it can be summarized as the cognitive mechanism responsible for active maintenance of and access to a limited amount of information crucial for the current task/action/thought (Cowan, 2017). WMC is typically measured using relatively simpler tasks (as compared to reasoning tests) that might require storage and recall of several items (simple span tasks), storage and recall under concurrent processing on the distracting task (complex-span tasks) or quick updating of the information maintained (memory updating tasks).

One potential way in which WM could constrain relation processing is the maximum number of slots, or amount of resources, available (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008). For instance, a person able to maintain only three elements in WM (a typical average capacity in the general population) could fail to process a quaternary relation, given an incapacity to grasp all its arguments at once. However, this possibility is not altogether plausible, as multiple studies on relation processing (e.g., Goldstone & Medin, 1994; Keane, 1997; Kubose, Holyoak, & Hummel, 2002) have suggested that even complex relations can be effectively processed in an incremental way (but see Halford et al., 2005). Another option of how WM could determine how well relations are processed is how strong is the control that can be exerted over WM contents. Strong control might block distractors (e.g., salient but invalid argument values) and remove elements that are no longer relevant, thereby facilitating relation processing (Jarosz & Wiley, 2012; Kane & Engle, 2002; Kovacs & Conway, 2016;). However, many studies failed to observe sufficiently strong correlations between hallmark executive control tasks and Gf (Chuderski, Taraday, Nęcka, & Smoleń, 2012; Colom et al., 2008; Friedman et al., 2006; Rey-Mermet, Gade, Souza, von Bastian, & Oberauer, 2019; Unsworth, Fukuda, Awh, & Vogel, 2014), with an exception for the antisaccades task (Rey-Mermet et al., 2019).

Alternatively, the authors of the relation monitoring task (Oberauer et al., 2007) have posed the *binding hypothesis*, assuming that the crucial building blocks underpinning relations in the mind consist of temporary bindings that are constructed, maintained, and finally dissolved in WM. Such bindings constitute "atomic" relations that can link one item to another (e.g., to make a pair) or an item with its context (e.g., to encode the item's position or membership). The binding construct seems ideally suited to explain the assignment of values to relational arguments (cf., Chuderski & Andrelczyk, 2015; Halford et al., 2007; Hummel & Holyoak, 1997). To process a relation (but also remember an ordered list or understand a longer utterance), one must be able to maintain in WM a sufficient number of bindings in parallel (Oberauer, 2019). In addition to upholding single bindings in WM, the ability to integrate such bindings into novel structural representations (so-called relational integration) can also be crucial. According to the binding hypothesis, effective relational integration enables the recall of a memory list in the proper order within the complex span, permits transforming several item-item bindings into noticing the row or column of identical symbols in the relation monitoring task, as well as poses a limit on relations that can be effectively processed in Gf tests. As a result, the capacity for bindings and relational integration is the driving force for the strong intercorrelations among all these kinds of tasks (Oberauer et al., 2008).

Exploring the role of temporary bindings further, a recent study (Chuderski, 2019) indicated that processing of even a single binding between two symbols, without needing to integrate them into a structure, can be crucial for Gf. In an apparently trivial task, only binary and meaningless relations between symbols needed to be compared (letters connected by "/" or "\" sign) to identify identical relations (i.e., that "A/B" is identical to "A/B" but not to "A\B"). Despite the fact that errors on the task were rare, their rate still predicted scores on RAPM and a figural analogies test surprisingly well (comparably to the two latter scores correlation). However, when the symbols were not bound by any sign and thus could be easily chunked into one representation ("AB"), the task no longer correlated with Gf. Thus, Gf may not only depend on the available quantity of bindings in WM but rather on their reliability and stability (a binding must prevail undistorted until the processing of relation is complete).

Summing up, there is a growing body of research suggesting that Gf primarily reflects how well one represents relations in the mind (even simple ones). However, the extent to which the construct of Gf can be translated into the ability to process relations has not been precisely defined. Consequently, the present study aimed to examine how far Gf, paradigmatically operationalized as the variance in scores on traditional fluid/relational reasoning tests—which require identifying unknown relations (in RAPM, discovering new rules); integrating them (combining up to five rules per a problem); inferring the missing values of their arguments (constructing the missing ninth cell), and —also most likely—involving learning (Bui & Birney, 2014; Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009); goal management (Carpenter, Just, & Shell, 1990); and various strategies (Jarosz et al., 2019)—can be explained in simpler and more univocal terms of processing a relation.

1.1. Relation processing tasks

To achieve this aim, we applied three relation processing tasks. Two tasks, *Monitoring* and *Comparing*, were modified variants of the previously-mentioned tasks developed by Oberauer et al. (2003) and Chuderski (2019), respectively. Also, we introduced a newly-developed task: *Matching* (see below). In all three tasks, the typical requirements of the Gf tests, mentioned above, are greatly reduced or absent. Each task requires validating whether the elements present on the screen do or do not satisfy a predefined relation. To achieve this, some relational information must be processed.

In *Monitoring*, participants need to validate the "all different" relation among several Greek letters in a row or column of a matrix (no two letters can be identical). Most likely, this requires rapid construction of temporary bindings among all the letter pairs in the given row/column and their integration into one relational representation. For example, four letters in a row/column (e.g., δ λ π θ) require building the maximum of six bindings (δ-λ, δ-π, δ-θ, λ-π, λ-θ, π-θ). The relation monitoring process for the given row/column stops when the identity between two letters is found (which falsifies the relation) or when no identity can be detected (and thus the relation becomes validated). We expect that the validation process of the "all different" relation is most effective when all the possible bindings among pairs of letters can be upheld and compared in WM in a single step. When not all such bindings can be held in a temporary store (due to limited capacity for bindings or attentional lapses), the process needs to be segmented and the letters bound and integrated into subgroups. In this case, the probability of missing an identity between a pair of letters increases and accuracy drops (for a process model of a variant of this task, see Chuderski, Andrelczyk, & Smolen, 2013).

In *Comparing*, participants need to decide whether two pairs of shapes (related to each other by either the "larger than" or" smaller than" relation) represent identical relations despite variant layouts of the shapes and the relation symbol (e.g., that ○ > □ means the same as □ < ○ but not the same as ○ < □ and ○ > Δ). We assume that to solve this task, a pair must first be represented as a proper binding of the objects to their specific roles in that relation (○-larger, □-smaller). Next, such a representation must be maintained in WM in order to be compared to other pairs (that also must be encoded in the same way). Some pairs may readily be excluded, e.g. when they include one or two shapes different from those in the maintained pair. Crucially, however, to correctly validate whether the two compared pairs that included the same shapes satisfy the same unique relation, the participant must maintain proper bindings between the specific shapes and their specific roles in that relation. In consequence, most errors in the task should result from misrepresenting such bindings or a failure in retaining them in WM. Therefore, the ability to properly encode and maintain stable relational representations during a concurrent processing should lead to more effective solving of the *Comparing* task.

Finally, in *Matching*, two perceptually-different but relationally-isomorphic directed graphs are presented. Participants must match two vertices across the two graphs. Each vertex can be identified by the number of incoming and outgoing edges (its degree). Encoding the degree requires representing a relation of the specific vertex with the number of its incoming and outgoing edges. Such a relation may be binary (e.g. "the vertex has one incoming edge") or may have more arguments (e.g. "one incoming and two outgoing edges"). Next, such a representation needs to be maintained in WM and compared to the degrees of other vertices, in order to validate whether the compared representations satisfy the same unique relation (e.g., "having one incoming and two outgoing edges"). Moreover, identification of the vertices and navigation through the problem elements is also achieved by utilizing the unique connections among the circles, which requires assigning the proper vertices into proper roles in the relation (e.g., "the vertex with one outgoing edge connects to the other specific vertex"). This applies especially to the most difficult graphs, in which participants need to discriminate between two vertices of the same degree by representing the pattern of their connections to other vertices. In consequence, errors in the task should result from misrepresenting the degrees of the vertices and the relations between two or more vertices.

1.2. Research questions

Our two main predictions were as follows: First, we expected that the variance in scores on relation processing tasks, which were designed to primarily require validation of whether the given arguments satisfy the predefined relation or not, would be *statistically equivalent* to Gf. If such variance could be equivalent to Gf, then most aspects of typical operationalization of Gf would not be necessary for measuring this construct (because the simple validation of whether the predefined relation holds or not for the given data unlikely involves relation discovery, inference, and learning; it probably also, to a degree, captures goal management and strategies). Therefore, such a result would purify the Gf construct. Moreover, a potential demonstration that the Gf variance equates variance in validation of the arguments of a relation can direct future research on the WM-based mechanisms of Gf (e.g., highlighting a potential role of temporary bindings). However, it must be noted that the present, purely correlational, study cannot allow for drawing conclusions about the exact (neuro)cognitive mechanisms underlying Gf and relation processing.

Second, on the basis of existing literature (Gignac, 2014; Kyllonen & Christal, 1990; Oberauer et al., 2005), we expected that WMC, as assessed with the typical simple span, complex span, and updating tasks, could not be statistically equivalent to Gf. We further predicted that our relation processing tasks would be able to explain the Gf variance (above and beyond WMC). On the one hand, the binding hypothesis predicts that the WM tasks, the relation processing tasks, and the Gf tests all rely on upholding items and bindings in WM—what likely drives their shared variance (see Oberauer et al., 2008). However, whereas the WM tasks primarily demand the holding of sheer items up to full memory capacity, the relation processing tasks and Gf tests typically do not (all the items are present on the screen or the paper sheet and can be attended to when needed). Moreover, both the Gf tests and the relation processing tasks require not only upholding the component bindings (the value-argument assignments) but also integrating them into a single complete relation (in the relation processing tasks) and even in a more complex structural representation (e.g., in RAPM and CFT-3), making their variances converge, whereas the WM tasks typically do not involve explicit relations (e.g., an unordered set of item-position bindings suffices to recall in the complex span task), causing WMC variance to diverge. In consequence, the capacity to uphold all the necessary information in WM (items, context, bindings), as reflected by high WMC, would not automatically lead to high Gf (even though the two would be strongly correlated), because even if all this information is present in WM, still the relation can be processed incorrectly (e.g., structure consistency can be violated), leading to relational errors.

To sum up, three relation processing tasks, three established reasoning tests, and three WMC tasks were administered. Each task category corresponded to their respective latent variable (factor): *Relation Processing*, *Fluid Reasoning*, and *WMC*. All three factors and their relationships were modeled using confirmatory factor analysis (CFA). The decisive model tested whether *Relation Processing* could be equivalent to *Fluid Reasoning* (i.e., if the correlation between these factors is statistically indistinguishable) while *Fluid Reasoning*, on the other hand, is distinguishable from *WMC*. Also, structural equation modeling (SEM) determined whether *Relation Processing* could account for Gf variance—above and beyond *WMC* (as well as fully mediate the link of *WMC* to *Fluid Reasoning*).

2. Method

2.1. Participants

Two hundred forty-five participants (154 women, 77 men) were

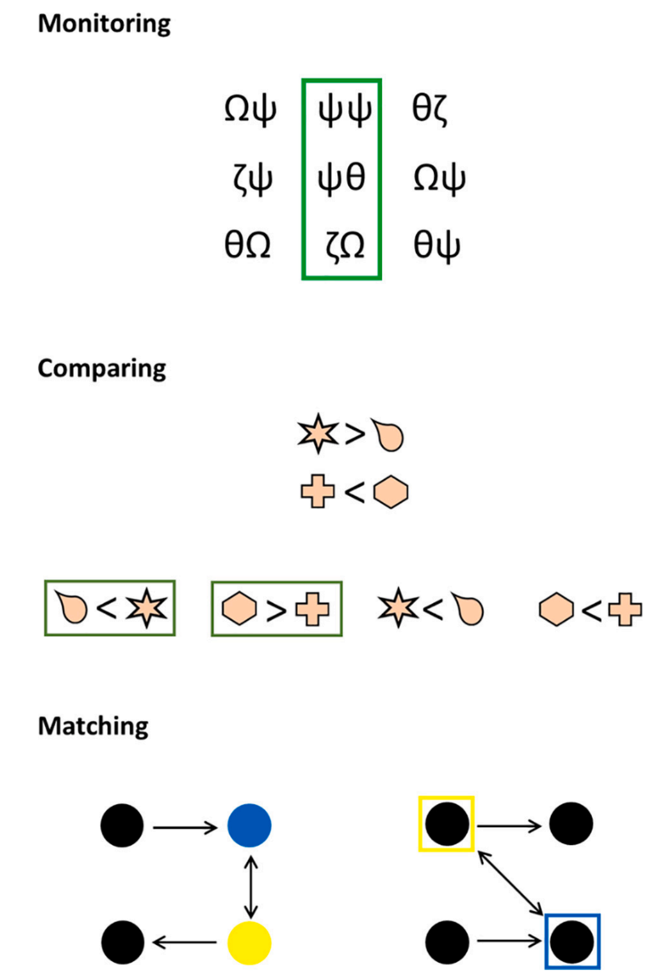

Fig. 1. Examples of trials from the three relation processing tasks applied in the study. The frames indicate the correct responses (not shown to participants).

recruited using ads on popular websites; participants were paid the equivalent of 10 euros for their participation. Their mean age was 22.8 years ($SD = 3.1$, range = 18–35). Most of the participants were students (56%). All had normal or corrected-to-normal vision. The participants were informed that participation was anonymous and that they could end their participation at will at any moment. The study fully conformed with the Declaration of Helsinki.

2.2. Relation processing tasks

For illustration of the tasks see Fig. 1. Each trial of *Monitoring* included an $n \times n$ array of pairs of Greek letters, presented at the center of the screen. The task required scanning the right-hand letters of the consecutive pairs, while the left-hand letters had to be ignored (their role was to prevent simple chunking of the right-hand letters). In all the pairs of each row and each column of the array, at least two right-hand letters were the same—except for either one row or one column (in which all the n right-hand letters were different). The goal was to find this unique row/column and mark it by clicking at any two of its pairs. The size of an array (n) increased from three up to six. The time limit for each of 50 trials was $7.0 + 2.5n$ s (i.e., 14.5–22.0 s), which suggested a time reserve left when compared to the mean RT of 6.29 s ($SD = 2.76$) for the easiest items, of 15.4 s ($SD = 5.61$) for the most difficult items, and the total mean RT of 8.87 s ($SD = 4.69$).

Each trial of *Comparing* consisted of n vertically-arranged pairs of simple geometric shapes bound by either the " > " or " < " sign (indicating either the "larger than" or "smaller than" relation between the two shapes, respectively). The number of types of shapes that appeared in each trial was $n + 2$. Thus, as the overall number of stimuli in the trial was $2n$, some figures belonged to more than one pair. The shapes were drawn from the pool of seven distinctive shapes. Below the vertical pairs, another four horizontally-arranged pairs were presented, which comprised the response bank. The goal was to select the two correct response options that expressed the same relation, as did some of the target pairs. Crucially, the order of shapes in each correct option was reversed relative to the respective target pair, for example, ○ > □ changed to □ < ○. Thus, the participant had to compare the meaning of each relation. In the incorrect options, the order of shapes was reversed relative to the corresponding pair while the sign did not change, or vice versa (i.e., the relation was opposite to the target relation). Only selecting the two correct options counted as the correct answer. The difficulty level (n) increased from two to five pairs throughout the task sequence. The time limit for each of 50 trials was $11.0 + 1.5n$ s (i.e., 14.0–18.5 s), while the mean RT equaled 7.21 s ($SD = 2.13$) for the easiest trials, 12.19 s ($SD = 3.53$) for the most difficult trials, and 10.60 s ($SD = 3.46$) in total.

In each trial of *Matching*, two directed graphs constructed out of circles (vertices) and arrows (edges) were presented adjacent to each other. The two graphs had the same structure of vertices and edges, but the location of the circles in the right-hand graph was transformed with respect to the left-hand graph. Always, two circles in the left-hand graph were highlighted, one in blue and one in yellow. The task was to find and mark these two circles in the right-hand graph by clicking on the corresponding circles with the left and the right mouse button for each color circle, respectively. A small mouse icon with the left and right button marked in the respective colors was always shown as a reminder. The blue and the yellow circle had always its exact counterpart in the right-hand graph that remained in the identical relation with all vertices and edges as was the case for the target circle. Matching correctly the two target circles across the graphs counted as the correct response. In most trials, each circle could be identified directly by its number of incoming and outgoing arrows (degree). In some potentially more difficult trials, the two targets had the same degree and therefore could only be discriminated indirectly, on the basis of the unique relation of the target with another circle. The number of circles in each graph increased from three in the easiest trials up to six in the most difficult trials. The number of edges varied between three and six. The time limit for each of 44 unique trials was 22 s. The mean RT equaled 10.54 s. ($SD = 3.85$) for the 10 easiest trials, 15.88 s. ($SD = 4.26$) for the 10 most difficult items, and 13.12 s. ($SD = 4.52$) in total.

Each task was preceded by directions that included two examples (with correct responses explained) and a practice session with several trials with feedback. Participants were advised that only accuracy on the tasks counted, not speed. To ensure that participants properly understood each task, the experimental trials started only if the accuracy on the practice trials reached the pre-specified satisfactory level. Participants had a maximum of five practice attempts to learn a task, but most of the participants needed just one attempt. In each task, a small clock icon was shown 5 s before the trial time limit, to advise the participant about the brief time remaining. The score on each task was the mean accuracy across all its trials.

2.3. Reasoning tests

The RAPM test (Raven et al., 1998) consists of 36 problems that include a 3×3 matrix of figural patterns, which is missing the bottom-right pattern, and eight response options comprising the patterns that could potentially match the missing one. The goal is to discover the rules that govern the distribution of patterns and apply them to the response options in order to choose the single correct pattern.

Three subtests (36 items in total) of CFT-3 (Cattell, 1961) were selected: Topology (all 10 items); Series (12 items; one problem was removed to match the desired number of problems: 36) and Classifications (14 items; the most difficult and virtually unsolvable problem was replaced with a newly-designed, easier one). Topology problems require choosing one option out of five, in which a dot matches all the topological relations to surrounding geometric elements as in the displayed example. The goal in Series problems is to find rules that govern the sequence of figures and to choose one out of six options that completes the sequence. Classification problems require choosing two figures out of five, that, according to some hidden rules, belong to the same artificial category, while the three remaining figures correspond to a different category. Each subtest had its separate instruction and time limit. The Matrices subtest was excluded from CFT-3, because it overlapped conceptually with RAPM.

The Figural Analogies Test (Analogies; Chuderski & Necka, 2012) consists of 36 analogies in the form of "A is to B as C is to X," where A, B, and C are relatively simple patterns of shapes. A is related to B according to two, three, four, or five rules (e.g., symmetry, rotation, change in size, color, thickness, number of objects), and X is an empty space. The task is to choose one pattern from an array of four which relates to C in a manner analogous to B to A.

Based on our previous data, each test and subtest was divided into two equipotent sets of items, both presenting a comparable average difficulty (i.e., mean accuracy). Such a division aimed at comparing the results between the typical paper-and-pencil administration of tests with their computerized administration, that is, we expected that correlations with the relation processing tasks would be the same regardless of the administration. The time limit for each set of RAPM, Analogies, Topology, Series, and Categories was 20, 15, 5, 6, and 7 min, respectively (these were standard administration times). Each set was preceded by instructions and two simple examples (including an explanation of the correct responses). The final score on each of the three tests was the average of correct responses in both sets (the scores on the three CFT-3 subtests were summed).

2.4. Working memory tasks

The spatial simple span task was modeled after a respective complex span task (Conway et al., 2005), but did not include the concurrent task. The task presented three, four, five, or six stimuli (set size) sequentially for 1.2 s each. Each stimulus was a 3 × 3 grid with one random cell filled red. After the grids presentation, participants had to recall the red filling locations in the proper order by pointing out with the mouse the respective cells on empty grids. The figural complex span task (also modeled after Conway et al., 2005) involved the same procedure, but the stimuli were shapes (e.g. squares, triangles) and, to prevent their chunking, each grid was followed by a concurrent task to decide whether a presented color is light or dark. Four and five trials per set size were applied in each task, respectively. A variant of mental counters task (Larson, Merritt, & Williams, 1988) presented a 120-item sequence of the A, B, and C letters in random order (with each letter displayed for 2.5 s.). The goal was to count how many times each letter had been presented, up to four times. Participants decided whether they had seen the letter either for the fourth time (target) or a lesser time than four (non-target). So, the letter counters had to be updated continuously. The final score was the hit rate for targets minus the false alarm rate for non-targets (as suggested by Snodgrass & Corwin, 1988).

2.5. Procedure

The study took place in a psychology laboratory at a central European university. Participants were tested in groups of two to eight people. First, two semantic analogy tests, which were part of a separate research project, were completed by participants; this took about 20 min. Next, the reasoning tests and the relation processing tasks were administered. Each participant completed two sets of RAPM, Analogies, and each of the CFT-3 subtests. Each set could be administered in either the paper-and-pencil format or on the computer. First, half the participants, randomly selected, completed the paper-and-pencil variants of the reasoning tests, whereas the other half completed their computerized variants. Next, the relation processing tasks was completed by all the participants. Then, the participants who had completed the paper-and-pencil variants of the reasoning tests attempted their computerized variants, whereas those who had completed the computerized variants attempted the paper-and-pencil variants. Finally, the three WM tasks were administered. The entire procedure took about 3.5 h and included a 10-min break. Snacks and drinks were provided to participants during the break.

2.6. Missing data and data preparation

Sixteen participants (6.5%) were excluded from all the following analyses because of lack of proper engagement, as indicated by very low scores and very short response times in at least one reasoning test or relation processing task, and/or because of missing multiple measures. Thus, the final dataset consisted of 229 participants. Three participants scored close to the guessing rate (< 0.20) on a single relation processing task but scored normally on all the remaining measures, so these single scores were removed, but the participants were not excluded. Two participants missed one set of a single reasoning test and seven participants missed one WM task. Due to a technical error, scores of as many as 35 participants on *Monitoring* were not saved. Also, two multivariate outliers were identified using Mahalanobis distance – the two outlying scores were removed but the participants were included. All the above missing values (2.2% of the dataset) were imputed by the expectation-maximization algorithm implemented in SPSS (IBM Corp, 2017).

3. Results

Table 1 shows descriptive statistics for all nine measures. The reliability of *Monitoring, Comparing,* and *Matching* was very good (McDonald's ωs around 90). All the measures approximated closely the normal distribution (skew < 2.0, kurtosis < 4.0). Table 2 shows the respective correlation matrix. The correlations of *Comparing* and *Matching* with reasoning tests were strong (from $r = 0.52$ to 0.61), while the correlations of *Monitoring* were slightly weaker ($r = 0.44$–0.55).

The key analysis consisted of the test of CFA Model 1a (Fig. 2) that included three factors: the latent variables of *Relation Processing, Fluid Reasoning,* and *WMC*. The fit of the model was good, $\chi^2(24) = 44.13$, Bentler's confirmatory fit index (CFI) = 0.977 (criterion value > 0.950; see Hu & Bentler, 1999), the root mean square error of approximation

Table 1
Descriptive statistics for the measures used in the study. N = 229.

Measure	Mean	*SD*	Min	Max	Skew	Kurtosis	ω
Monitoring	0.80	0.14	0.22	1.00	−1.6	3.58	0.90
Comparing	0.75	0.17	0.28	0.98	−0.9	−0.05	0.89
Matching	0.65	0.16	0.23	0.98	−0.3	−0.34	0.95
RAPM (total score)	0.65	0.15	0.25	0.97	−0.2	−0.36	0.83
Analogies (total score)	0.71	0.14	0.31	1.00	−0.5	−0.13	0.77
CFT-3 (total score)	0.64	0.13	0.28	0.94	−0.4	−0.17	0.73
RAPM – paper	0.66	0.17	0.22	0.94	−0.4	−0.33	–
Analogies – paper	0.73	0.16	0.22	1.00	−0.7	0.23	–
CFT-3 – paper	0.67	0.15	0.17	0.94	−0.5	0.01	–
RAPM – computer	0.63	0.17	0.11	1.00	−0.2	−0.24	–
Analogies – computer	0.69	0.16	0.22	1.00	−0.5	0.37	–
CFT-3 – computer	0.62	0.14	0.17	1.00	−0.4	0.17	–
Simple span	0.80	0.11	0.37	1.00	−0.4	0.19	0.80
Complex span	0.58	0.17	0.17	0.97	−0.2	−0.40	0.86
Mental counters	0.51	0.16	0.00	0.96	0.0	0.33	0.93

Table 2
The correlation matrix. All ps < 0.005. N = 229.

Measure		1.	2.	3.	4.	5.	6.	7.	8.
1.	Monitoring	–							
2.	Comparing	0.59	–						
3.	Matching	0.44	0.57	–					
4.	RAPM	0.55	0.56	0.60	–				
5.	Analogies	0.50	0.60	0.61	0.68	–			
6.	CFT-3	0.44	0.52	0.61	0.66	0.58	–		
7.	Simple span	0.35	0.42	0.51	0.53	0.41	0.43	–	
8.	Complex span	0.28	0.25	0.33	0.38	0.37	0.34	0.36	–
9.	Mental counters	0.29	0.36	0.39	0.41	0.35	0.35	0.35	0.29

(RMSEA) = 0.056 [0.024, 0.084] (criterion value < 0.060). The factor loadings for *Comparing* and *Matching* were high (0.72 and 0.76), with a slightly lower but still satisfactory loading for *Monitoring* (0.64; criterion value > 0.40; Stevens, 1992). The crucial correlation between *Fluid Reasoning* and *Relation Processing* was very strong ($r = 0.950$), as well, a test of the difference between two dependent correlations (Lee & Preacher, 2013) showed that this correlation was significantly stronger than the *WMC-Fluid reasoning* link ($r = 0.857$), z = 8.06, $p < .001$.

Next, in Model 1b, the *Relation Processing-Fluid Reasoning* link was constrained to unity. The change in fit was nonsignificant, $\Delta\chi^2(2) = 3.38$, $p = .185$, indicating that *Relation Processing* may be considered statistically equivalent to *Fluid Reasoning*. At the same time, constraining the *Fluid Reasoning-WMC* link to unity in Model 1c led to a significant change in fit, $\Delta\chi^2(2) = 7.52$, $p = .023$, so *WMC* was statistically distinct from *Fluid Reasoning*.

Finally, Model 1a was compared to the two-factor Model 2 in which *Fluid Reasoning* loaded both the three reasoning tests and the three relation processing tasks; the second factor was *WMC*. The difference in fit between the models was negligible, as indicated by Akaike's information criterion (AIC; with lower value indicating better fit), $AIC_{Model1a} = 86.13$, $AIC_{Model2} = 85.51$, $\Delta AIC = 0.62$. The fit of a model in which all the nine measures loaded a single factor was worse, AIC = 90.62, ΔAIC = 4.49, again supporting a relative statistical distinctness of *WMC*.

To control for the potential influence of the reasoning test format on the correlation with relation processing tasks (which were all administered on computer), we tested CFA Model 3, which included separate factors for each reasoning test format (paper vs. computer), as well as *Relation Processing* and *WMC*. The model fitted well, $\chi^2(48) = 83.93$, CFI = 0.968, RMSEA = 0.053 [0.031, 0.074]. As expected, the format did not affect validity of the reasoning tests: the two alternative factors of *Fluid Reasoning* were identical ($r = 1.0$). Also, the strength of correlation with *Relation Processing* ($r = 0.955$ and $r = 0.921$) and *WMC* ($r = 0.837$ and $r = 0.857$) was very similar for the paper and computer formats, respectively, and was close to these two factors' correlations with the full *Fluid Reasoning* factor in Model 1.

Finally, SEM Model 4a (Fig. 3a) tested whether *Relation Processing* could explain *Fluid Reasoning* beyond the variance explained by *WMC* (the fit of the model was identical to that of Model 1a). In this model, *WMC* predicted *Fluid Reasoning* and *Relation Processing*; the disturbance terms of these two variables, representing variance not explained by the predictor, were correlated. Their correlation ($r = 0.81$) was substantial, demonstrating that most of the variance in relation processing tasks and reasoning tests that could not be explained by *WMC* was shared between these two variables. In line with this result, the SEM Model 3b (Fig. 4b) showed that *Relation Processing* fully mediated the *WMC-Fluid Reasoning* link, as the latter path became nonsignificant in this model when *Relation Processing* predicted *Fluid Reasoning*. This result suggested that the specific *WMC* variance, not shared with *Relation Processing*, failed to predict *Fluid Reasoning*. At the same, *WMC* did not mediate the *Relation Processing-Fluid Reasoning* link.

One limitation of the above results is that the three relation processing tasks did not contribute equally to the actually perfect correlation between the *Fluid Reasoning* and *Relation Processing* latent variables. Fig. 4 shows the relationship of each task score with the *Fluid Reasoning* factor *z*-score (Gf), as well as each task's score distribution. The lower factor loadings and correlation coefficients observed for *Comparing*, and especially *Monitoring*, as compared to *Matching*, could to some extent result from the imperfect distributions and slightly limited ranges due to the certain ceiling effect. Still, the strength of the Gf-*Monitoring* and Gf-*Comparing* correlations was similar to the ones previously reported in the literature for the similar variants of these tasks ($r \approx 0.5$–0.6; Bateman et al., 2019; Chuderski, 2019; Oberauer et al., 2008). All these zero-order correlations were also stronger than that usually observed for single measures of *WMC* ($r \approx 0.3$–0.5; e.g., Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth et al., 2014). In line with this, the correlations of the simple span, complex span, and mental counters tasks with Gf amounted to $r = 0.52$, 0.42, and 0.42, respectively, suggesting an inferior predicting power of *WMC*—even at the level of single tasks.

4. Discussion

Three tasks that required validating a simple relation—that several symbols were mutually distinct, that one element was superior over the other (irrespective of their spatial order) and, finally, that pairs of nodes were interconnected in two graphs in a corresponding way—appeared to require virtually the same ability tapped by the much more complex fluid reasoning tests. The latent variable composed of three relation processing tasks and the latent variable reflecting Gf were statistically equivalent regardless of whether the latter was measured on the computer or using paper and pencil. The relation processing factor shared with Gf as much as 90% of the variance, above and beyond the explanatory power of WMC (73%). The latter variable could be clearly distinguished from Gf. Finally, the relation processing tasks mediated

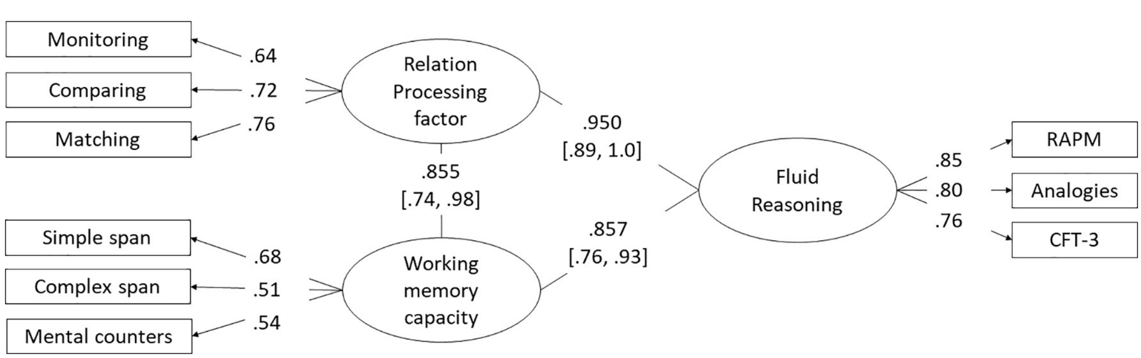

Fig. 2. CFA Model 1a showing an almost perfect correlation between Relation Processing and Fluid Reasoning. The ovals represent factors and the rectangles represent manifest variables. The values on the arrows are standardized factor loadings and the values between the ovals are correlation coefficients with 95% confidence intervals.

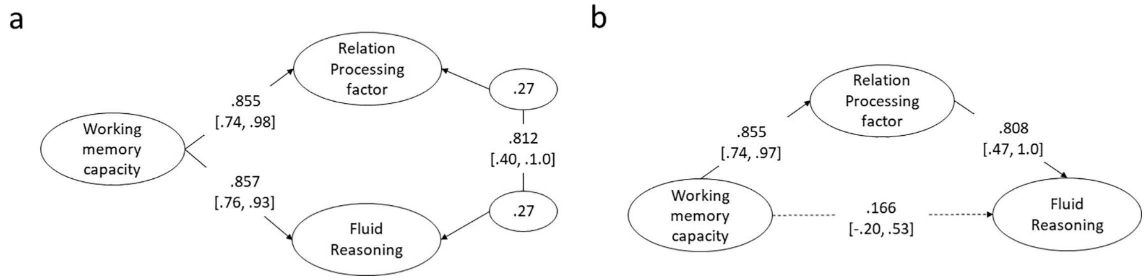

Fig. 3. (a) SEM Model 4a, in which WMC predicts Fluid Reasoning and Relation Processing factors. The numbers inside the small ovals are disturbance terms reflecting the amount of variance not explained by the predictor. The value between the small ovals is the correlation between the disturbance terms. (b) SEM Model 4b in which Relation Processing fully mediates the WMC-Fluid Reasoning link. The dotted line indicates a nonsignificant path.

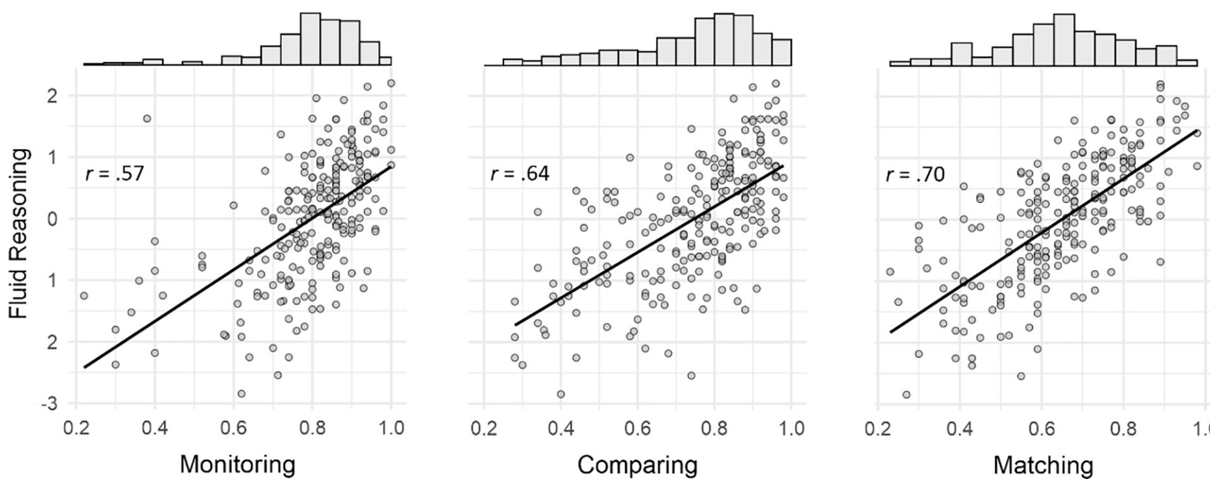

Fig. 4. Scatterplots showing the relationship between accuracy on each of three relation processing tasks and the z-score on Fluid Reasoning factor (Gf) calculated from three reasoning tests. The histograms above the scatterplots show the distribution of each task. N = 229.

the relationship between WMC and Gf, sharing with Gf as much as 66% of the residual variance left unexplained by WMC. These results are consistent with previous work that has reported strong correlations between Gf and various relational processing tasks (e.g., Bateman et al., 2019; Birney & Bowman, 2009; Buehner, Krumm, & Pick, 2005; Chuderski, 2014, 2019; Krumm et al., 2009; Oberauer et al., 2008; Wilhelm et al., 2013).

Therefore, these results allow for purifying the Gf construct, by showing that some complex cognitive operations, typically evoked in definitions of Gf (see McGrew, 2009), are in fact dispensable in measuring fluid reasoning. First, in the relation processing tasks, the type of relation was defined a priori and the relation was provided explicitly (the participants were even pre-trained on using the relation). Thus, the processes of discovering the relation and learning do not seem crucial for Gf. Second, as each relation processing task involved a simple single relation, Gf is not about the complex integration of multiple relations. Finally, as the straightforward instructions for these tasks left little room for factors other than encoding stimuli, constructing bindings, and/or representing and applying relations, differences in such factors were unlikely to substantially impact the scores.

These results should not be interpreted as suggesting solely that the relation processing tasks provide a more parsimonious alternative to the existing fluid reasoning tests. Rather, these results have crucial theoretical ramifications. The fact that the relatively simple and univocal tasks explained almost entire variance in fluid reasoning indicates that the latter ability (to date not defined sufficiently explicitly) is actually equivalent to the effectiveness in constructing and maintaining a single relational representation in the mind (e.g., "$\Psi \neq \Phi$ & $\Phi \neq \Omega$ & $\Psi \neq \Omega$" in *Monitoring* or "$\lozenge < \Delta$" in *Comparing*) which serves to determine if a given option satisfies the relation represented (i.e., "all the three symbols differ," or "$\Delta > \lozenge$"), or not.

In a way, the present results invoke Spearman's (1927) claims that the mental ease of dealing with relations is fundamental to fluid intelligence (we note that Spearman actually wrote about the *g* factor but, in fact, his conceptualization of *g* makes it quite close to Gf; see Gustafsson, 1984; McGrew, 2009). However, our results do not simply confirm Spearman's (1927) (kind of vague) hypothesis but they rather suggest which aspects of relation processing matter for Gf. The role of educing relations, i.e., "[the] power to bring to mind any relations that essentially hold between two or more ideas" (p. 165), as noted above, does not seem necessary, as relation processing is practically equivalent to Gf—even if the relation in question is revealed to the participants (it need not "be brought to mind" per se). The role of educing correlates, i.e., "[having] in mind any idea together with a relation, . . . power to bring up to mind the correlative idea" (p. 166), does not seem necessary either. Perhaps only in the *Matching* task, projecting the structure of the left-hand graph onto the right-hand graph—"making the mapping" in modern terminology—can be interpreted as Spearman's "bringing the correlative idea," given that their correlation stems from a shared relational structure. However, the two other relation processing tasks barely involve processes that could be interpreted as educing correlates (see Section 1.1). Crucially, the present results suggest that individual differences in cognitive ability may be driven primarily by "validation of correlates": checking whether the current assignment of values to the consecutive arguments does, or does not, satisfy the target relation. In consequence, Spearman was right that Gf primarily reflects the ability to deal with relations; however, the crux of defining Gf in terms of relations seems to be simpler than Spearman had envisioned.

Furthermore, although this study cannot shed light on the precise cognitive mechanisms underlying relation processing and Gf, our second research question referred to WMC as the key factor likely driving the shared variance among WM tasks, relation processing tasks, and reasoning tests. WMC provides the necessary resources to perform each of the three categories of tasks, as all of them rely on upholding items and constructing and maintaining bindings in WM. Consequently, as expected, WMC explained most of the variance in relation processing tasks and reasoning tests. However, while WM tasks may require maintaining separate bindings (for example between items and their temporal or spatial order), they do not involve processing explicit relations. Therefore, we think that the shared higher-level requirement of relation processing tasks and reasoning tests that cannot be accounted for by WMC is the need to properly integrate the bindings in order to process a complete relation. Even if all the component information is properly represented in WM, the invalid assignment of items to the arguments of a given relation will result in a relational error. Thus, we think that the ability which allows to effectively validate relations in the relation processing tasks as well as enables processing more complex structural representations in the reasoning tests is critical for Gf, while it cannot be simply reduced to WMC.

Although the present study did not measure other elementary cognitive processes besides WMC, such as vigilance, attention control, short-term storage, long-term memory, processing speed etc., so their contributions to Gf could not be compared directly with the contribution of relational processing, to date WMC has been the strongest predictor of Gf (with correlations ranging 0.60–0.80), while the former contributions typically fell far below $r = 0.60$ (e.g., Chuderski et al., 2012; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Kane et al., 2004; Unsworth & Spillers, 2010). Therefore, as in the present study, WMC contributed to Gf less strongly than did relation processing, it is unlikely that the latter – virtually full – contribution could be surpassed by any other elementary process.

The present correlational study contributes to intelligence research by improving and clarifying the understanding of the Gf construct and by indicating a new manner of operationalization. This may in turn help to inspire future correlational as well as experimental studies that could analyze in greater depth, the relation-processing operations underlying Gf. In future studies, the more precise experimental control over task demands is needed to better isolate and possibly disentangle such abstract operations as relation validation and other components of relational reasoning (for such a recent attempt, see Shokri-Kojori & Krawczyk, 2018). The relative simplicity of the relation processing tasks makes them suitable for this purpose. Moreover, future studies might combine the relation processing tasks with both cognitive experimentation and brain imaging, to shed light on the precise neurocognitive mechanisms underlying relation processing. Finally, representing and applying relations in the mind/brain should be formalized in models that would explain these operations in terms of the underlying mechanisms and elaborate the source of constraints and errors in human relation processing (for examples of such attempts, see Chuderski & Andrelczyk, 2015; Doumas, Hummel, & Sandhofer, 2008; Rasmussen & Eliasmith, 2014; Wilson, Halford, Gray, & Phillips, 2001).

Apart from its theoretical contribution, the present study also yields practical implications for intelligence research in that it validates relation processing tasks as useful Gf tests. These tasks may overcome some of the disadvantages of the traditional reasoning tests. Firstly, reasoning takes time: a typical reasoning test, on average, requires approximately 1 min for an item to be solved (40 min is recommended for the 36 RAPM items). Administration of three different reasoning tests, necessary to compute the latent Gf variable, may last up to one hour, even in the case of abbreviated test variants (note that decreasing the allotted time may produce unwanted effects of time pressure; see Chuderski et al., 2013; Estrada, Román, Abad, & Colom, 2017; Ren, Wang, Sun, Deng, & Schweizer, 2018). The relation processing tasks have significantly shorter completion times—although each task had a comparable number of items as a full reasoning test, each task could be completed in 10 min on average. Another improvement present in the relation processing tasks is a decreased guessing probability, resulting from the requirement to double select the items that make the correct response. Together with ensuring (via the training session) the proper understanding of instructions for the tasks, this might have resulted in the tasks' high reliability (≈ 0.90). Finally, as people become increasingly familiar with typical Gf tests, such as matrix or series completion, the cultural fairness of these tests and their validity may be corrupted (and, with that, the very idea of Gf as dealing with novel, knowledge-lean problems could be threatened). Instead, the relation processing tasks are highly unfamiliar for each solver. Also, as the tasks do not resemble typical reasoning tests, the solvers may be oblivious to their real purpose (i.e., testing individual differences in reasoning ability). Thus, the relation processing tasks proposed in this investigation can be an effective psychological tool in basic research and in many applications.

In conclusion, in the current study, two existing relation processing tasks were adopted and one new task was developed, to measure the effectiveness of validation of a simple single relation at the factor level. The latent variable loading of the three relation processing tasks explained almost the entire variance in Gf, as measured with typical reasoning tests. It also outperformed WMC in predicting Gf, sharing with Gf most of the residual variance left unexplained by WMC. These results enable greater clarification and better understanding of the Gf construct, showing what is crucial to its comprehension. The results indicate that many complex requirements (including rule discovery and integrations) are not necessary to capture the full variance in Gf. Instead, the ability to reason in novel abstract problems may be equivalent to something much simpler: the ability to effectively validate a single predefined relation.

Declaration of Competing Interest

None.

Acknowledgements

Data and materials can be downloaded from https://osf.io/v2ydz/ or received upon request. The study was supported by National Science Centre of Poland (grant #2019/33/B/HS6/00321).

References

Bateman, J. E., Thompson, K. A., & Birney, D. P. (2019). Validating the relation-monitoring task as a measure of relational integration and predictor of fluid intelligence. *Memory and Cognition*. https://doi.org/10.3758/s13421-019-00952-2.

Birney, D., & Bowman, D. B. (2009). An experimental-differential investigation of cognitive complexity. *Psychology Science Quarterly, 51*(4), 449–469. Retrieved from http://p16277.typo3server.info/fileadmin/download/PschologyScience/4-2009/psq_4_2009_449-469.pdf.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning = working memory ≠ attention. *Intelligence*. https://doi.org/10.1016/j.intell.2005.01.002.

Bui, M., & Birney, D. P. (2014). Learning and individual differences in Gf processes and Raven's. *Learning and Individual Differences, 32*, 104–113. https://doi.org/10.1016/j.lindif.2014.03.008.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review, 97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404.

Carroll, J. B. (1993). *Human cognitive abilities.* Cambridge: Cambridge University Press.

Cattell, R. B. (1961). *Culture free intelligence test, scale 3.* Champaign, IL: Institute for Personality and Ability Testing.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*. https://doi.org/10.1037/h0046743.

Chuderski, A. (2014). The relational integration task explains fluid reasoning above and beyond other working memory tasks. *Memory & Cognition, 42*(3), 448–463. https://doi.org/10.3758/s13421-013-0366-x.

Chuderski, A. (2019). Even a single trivial binding of information is critical for fluid intelligence. *Intelligence, 77*(July), 101396. https://doi.org/10.1016/j.intell.2019.101396.

Chuderski, A., & Andrelczyk, K. (2015). From neural oscillations to reasoning ability: Simulating the effect of the theta-to-gamma cycle length ratio on individual scores in a figural analogy test. *Cognitive Psychology, 76*, 78–102. https://doi.org/10.1016/j.cogpsych.2015.01.001.

Chuderski, A., Andrelczyk, K., & Smolen, T. (2013). An oscillatory model of individual differences in working memory capacity and relational integration. *Cognitive Systems Research, 24*, 87–95. https://doi.org/10.1016/j.cogsys.2012.12.005.

Chuderski, A., & Necka, E. (2012). The contribution of working memory to fluid reasoning: Capacity, control, or both? *Journal of Experimental Psychology, Learning, Memory, and Cognition, 38*(6), 1689–1710. https://doi.org/10.1037/a0028465.
Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence, 40*(3), 278–295. https://doi.org/10.1016/j.intell.2012.02.010.
Colom, R., Abad, F. J., Quiroga, M.Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*. https://doi.org/10.1016/j.intell.2008.01.002.
Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.
Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196772.
Cowan, N. (2017). Working memory: The information you are now thinking of. In J. T. Wixted, & J. H. Byrne (Eds.). *Cognitive Psychology of Memory, Vol. 2 of Learning and Memory: A Comprehensive Reference* (pp. 147–161). (2nd ed.). Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21040-7.
Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, 63*(1), 453–482. https://doi.org/10.1146/annurev-psych-120710-100353.
Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1–43.
Doumas, L. A. A., & Hummel, J. E. (2010). Computational models of higher cognition. In K. J. Holyoak, & R. G. Morrison (Eds.). *The Oxford Handbook of Thinking and Reasoning*(pp. 52–66). New York, NY: Oxford University Press.
Estrada, E., Román, F. J., Abad, F. J., & Colom, R. (2017). Separating power and speed components of standardized intelligence measures. *Intelligence, 61*, 159–168. https://doi.org/10.1016/j.intell.2017.02.002\.
Friedman, N. P., Miyake, A., Corley, R., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are, related to intelligence. *Psychological Science, 17*(2), 172–179. https://doi.org/10.1111/j.1467–9280.2006.01681.x.
Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.). *Categorization inside and outside the lab* (pp. 151–175). Washington, DC: APA.
Gignac, G. E. (2014). Fluid intelligence shares closer to 60% of its variance with working memory capacity and is a better indicator of general intelligence. *Intelligence, 47*, 122–133. https://doi.org/10.1016/j.intell.2014.09.004.
Goodwin, G. P., & Johnson-Laird, P. N. (2008). Transitive and pseudo-transitive inferences. *Cognition, 108*(2), 320–352. https://doi.org/10.1016/j.cognition.2008.02.010.
Goldstone, R. L., & Medin, D. L. (1994). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 29–50.
Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*. https://doi.org/10.1016/0160-2896(84)90008-4.
Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(6), 803–864. https://doi.org/10.1017/S0140525X98001769.
Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science, 16*(1), 70–76. https://doi.org/10.1111/j.0956-7976.2005.00782.x.
Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences, 11*, 236–242. https://doi.org/10.1016/j.tics.2007.04.001.
Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's advanced progressive matrices. *Journal of Vision*. https://doi.org/10.1167/11.10.10.
Holyoak, K. J. (2012). Analogy and relational reasoning. *The Oxford handbook of thinking and reasoning*https://doi.org/10.1093/oxfordhb/9780199734689.013.0013.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. https://doi.org/10.1080/10705519909540118.
Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*. https://doi.org/10.1037/0033-295X.104.3.427.
IBM Corp (2017). *IBM SPSS statistics for windows, version 25.0.* Armonk, NY: IBM Corp.
Jarosz, A. F., Raden, M. J., & Wiley, J. (2019). Working memory capacity and strategy use on the RAPM. *Intelligence*. https://doi.org/10.1016/j.intell.2019.101387.
Jarosz, A. F., & Wiley, J. (2012). Why does working memory capacity predict RAPM performance? A possible role of distraction. *Intelligence, 40*(5), 427–438. https://doi.org/10.1016/j.intell.2012.06.001.
Johnson-Laird, P. N. (2012). Inference with mental models. *The Oxford handbook of thinking and reasoning*https://doi.org/10.1093/oxfordhb/9780199734689.013.0009.
Kane, M. J., & Engle, M. J. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*, 637–671.
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology. General, 133*, 189–217.
Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66.
Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence, 37*(4), 374–382. https://doi.org/10.1016/j.intell.2009.03.004.
Keane, M. T. (1997). What makes an analogy difficult? The effects of order and causal structure on analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 946–967.
Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946.
Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37*(4), 347–364. https://doi.org/10.1016/j.intell.2009.02.003.
Kubose, T. T., Holyoak, K. J., & Hummel, J. E. (2002). The role of textual coherence in incremental analogical mapping. *Journal of Memory and Language, 47*, 407–435.
Kyllonen, P., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity ?!. *Intelligence, 433*, 389–433.
Larson, G. E., Merritt, C. R., & Williams, S. E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence*. https://doi.org/10.1016/0160-2896(88)90012-8.
Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.
McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*. https://doi.org/10.1016/j.intell.2008.08.004.
Oberauer, K. (2019). Working memory capacity limits memory for bindings. *Journal of Cognition, 2*(1), 40. https://doi.org/10.5334/joc.86.
Oberauer, K., Su, H., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory—storage processing supervision and coordination. *Intelligence, 31*, 167–193.
Oberauer, K., Süβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence, 36*(6), 641–652. https://doi.org/10.1016/j.intell.2008.01.007.
Oberauer, K., Wilhelm, O., Schulze, R., & Süß, H. M. (2005). Working memory and intelligence -their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.131.1.61.
Oberauer, K., Süß, H. M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.). *Variation in working memory*Oxford: Oxford University Press.
Rasmussen, D., & Eliasmith, C. (2014). A spiking neural model applied to the study of human performance and cognitive decline on Raven's advanced progressive matrices. *Intelligence*. https://doi.org/10.1016/j.intell.2013.10.003.
Raven, J., Raven, J. C., & Court, J. (1998). *Manual for Raven's progressive matrices and vocabulary scales.* https://doi.org/10.1006/cogp.1999.0735.
Raven, J. C. (1938). *Progressive matrices.* London: Lewis.
Ren, X., Wang, T., Sun, S., Deng, M., & Schweizer, K. (2018). Speeded testing in the assessment of intelligence gives rise to a speed factor. *Intelligence*. https://doi.org/10.1016/j.intell.2017.11.004.
Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General, 148*(8), 1335–1372. https://doi.org/10.1037/xge0000593.
Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*, 116–141. https://doi.org/10.1016/j.jml.2014.01.004.
Shokri-Kojori, E., & Krawczyk, D. C. (2018). Signatures of multiple processes contributing to fluid reasoning performance. *Intelligence, 68*(March), 87–99. https://doi.org/10.1016/j.intell.2018.03.004.
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34–50. https://doi.org/10.1037/0096-3445.117.1.34.
Snow, R., Kyllonen, P., & Marshalek, B. (1984). The topography of ability and learning correlations. *Advances in the Psychology of Human Intelligence, 2*, 47–103.
Spearman, C. (1927). *The abilities of man.* New York: Macmillan.
Stevens, J. P. (1992). *Applied Multivariate Statistics for the Social Sciences (2nd edition).* Hillsdale, NJ: Erlbaum.
Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71C*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003.
Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*(4), 392–406. https://doi.org/10.1016/j.jml.2010.02.001.
Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in psychology, 4*, 433. https://doi.org/10.3389/fpsyg.2013.00433.
Wilson, W. H., Halford, G. S., Gray, B., & Phillips, S. (2001). The STAR-2 model for mapping hierarchically structured analogs. In D. Gentner, K. J. Holyoak, & B. Kokinov (Eds.). *The analogical mind: Perspectives from cognitive science* (pp. 125–159). Cambridge, MA: MIT.